

VU Research Portal

Aligning Large SKOS-Like Vocabularies: Two Case Studies

Tordai, A.; van Ossenbruggen, J.R.; Schreiber, G.; Wielinga, B.J.

published in

The Semantic Web: Research and Applications
2010

DOI (link to publisher)

[10.1007/978-3-642-13486-9_14](https://doi.org/10.1007/978-3-642-13486-9_14)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Tordai, A., van Ossenbruggen, J. R., Schreiber, G., & Wielinga, B. J. (2010). Aligning Large SKOS-Like Vocabularies: Two Case Studies. In L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, & T. Tudorache (Eds.), *The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 –June 3, 2010, Proceedings* (Vol. 1, pp. 198-212). (Lecture Notes in Computer Science; Vol. 6088). Springer Verlag. https://doi.org/10.1007/978-3-642-13486-9_14

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Aligning Large SKOS-like vocabularies: Two Case Studies

Anna Tordai, Jacco van Ossenbruggen, Guus Schreiber and Bob Wielinga

VU University Amsterdam, Amsterdam The Netherlands

Abstract. In this paper we build on our methodology for combining and selecting alignment techniques for vocabularies, with two alignment case studies of large vocabularies in two languages. Firstly, we analyze the vocabularies and based on that analysis choose our alignment techniques. Secondly, we test our hypothesis based on earlier work that first generating alignments using simple lexical alignment techniques, followed by a separate disambiguation of alignments performs best in terms of precision and recall. The experimental results show, for example, that this combination of techniques provides an estimated precision of 0.7 for a sample of the 12,725 concepts for which alignments were generated (of the total 27,992 concepts). Thirdly, we explain our results in light of the characteristics of the vocabularies and discuss their impact on the alignments techniques.

Key words: vocabulary alignment. case study, cultural heritage, methodology

1 Introduction

As Semantic Web technology gains prevalence the field of ontology alignment is becoming more and more important. Within the MultimediaN E-Culture project [11] we use a large number of vocabularies for the annotation of artwork metadata. Despite the large amount of work done on developing ontology alignment techniques [7], in a practical setting it is still hard to predict, for two given vocabularies, which combination of techniques can best be used to create an alignment between them.

In previous work [12] we took a first step towards a methodology for selecting such a combination. We applied three alignment techniques to two vocabularies from the E-Culture repository, and looked which combination of techniques gave the best results. In this study, we take a second step by questioning to what extent we can use an analysis of the characteristics of the vocabularies to predict the performance of the different techniques, and to predict which combination will generate the best results. To answer this question, we perform two case studies, in each we align a large domain-specific vocabulary to a lexical resource. We analyze the vocabularies being aligned and predict the performance of four selected alignment techniques. We then apply all four techniques to generate alignments, followed by manual evaluation of representative samples to assess

the performance of each technique. Finally, we discuss our findings and compare them to our initial predictions.

2 Related Work

Work on procedures and guidelines for ontology and vocabulary alignment is still limited. The Ontology Alignment Evaluation Initiative(OAEI¹) campaigns provide a standardized way of comparing alignment tools, with tools such as Falcon [10] having the best performance in the 2007 [6] and RiMOM [13] among the top performers in 2008 [2] and 2009 [5] campaigns. Unfortunately, there are no clear selection criteria for these tools and many of the off-the-shelf tools are either unavailable or do not work on data other than that of the OAEI campaigns. Euzenat et al. [4] identify applications requirements and propose a case-based method for recommending alignment techniques but the work remains at a high level of abstraction. In a survey of alignment techniques, Aleksovski et al. [1] list techniques for alignment problems based on real world applications. They recommend using those techniques for similar application. In both cases there is a lack of a systematic method of comparison and evaluation of techniques with respect to domains and vocabulary characteristics. Eckert et. al. [3] used machine learning techniques for alignment generation but found that combining the results of multiple alignment tools by a system of voting works just as well as machine learning. This result suggests that machine learning techniques, while useful in other areas such as natural language processing, have little added value in the field of ontology alignment. Ghazvinian et. al. [8] compare the performance of off-the-shelf alignment tool and a simple lexical algorithm for creating alignments between medical ontologies using the OAEI gold standard. They concluded that the simple lexical algorithm performs better than the alignment tool.

Our conclusion is that we need to develop our own methods for aligning vocabularies with clear selection criteria.

3 Alignment Techniques

In previous work we learned that for vocabularies containing many synonyms or alternative labels, simple string matching techniques can already yield relatively good results at low computational costs. In contrast, off-the-shelf structural tools such as Falcon tend to find few or no extra alignments, but are computationally so expensive that they cannot be run on large vocabularies. For this study, we tested Falcon on our data set but it ran out of memory. A test run on a relatively small subset of a source vocabulary required 15 Gb of memory and 46 hours of runtime to generate alignments. Ghazvinian et. al. [8] reported similar problems with Falcon, and other off-the-shelf tools. In this paper we focus on relatively simple alignment techniques based on string matching. Variations within these techniques, however, still have a significant influence on the quality

¹ <http://oei.ontologymatching.org/>

of the alignments generated. Our goal is to be able to predict which variation will perform best, based on an analysis of the characteristics of the vocabularies being aligned.

3.1 Vocabulary characteristics

We focus on vocabularies that are either represented directly in SKOS or can be easily mapped to the SKOS model. Since relatively few organisations publish their vocabularies in this format, this typically means the SKOS version is the result of a conversion from some other format. Both the conventions during the development of the original format and the decisions made during the conversion influence the characteristics of the resulting SKOS vocabularies, especially when the source and target vocabulary follow different conventions. This is typically a source of potential alignment problems.

Consider the use of plural versus singular nouns forms in labels. Lexical resources tend to prefer singular, while the ISO standard for thesauri² prescribes the use of the plural form, and yet other vocabularies contain both forms.

Another example is the use of preferred and alternative labels versus the use of synonyms of equivalent status. The first is common practice in many domain-specific thesauri, the latter is commonly found in dictionaries and other lexical resources.

Another potential source of matching problems are the spelling conventions of words with upper case characters, diacritics and hyphens. For example, “Fin-de-siècle” may be spelled in this way in one vocabulary but as “fin de siecle” in another.

Finally, vocabularies tend to differ in the treatment of homonyms, that is, terms with the same label that have different meanings. Some vocabularies prevent homonyms by explicitly adding qualifiers to labels so that each label is unique. Others allow multiple concepts to have the same label, and rely on the concept’s place in the hierarchy or its scope note to clarify its meaning.

For all the examples given above, it is *a priori* not clear how the different conventions should be handled during the alignment and how this might influence the results.

3.2 Alignment Generation Techniques

In this paper we thus focus on three morphological techniques based on simple string matching, using either exact string matching, string matching after normalization of hyphens and diacritics, and string matching after conversion of plurals to singular form. We also look at the effect of using only preferred labels versus the use of both preferred and alternative labels. Application of these techniques yields a single alignment candidate for some concepts, but due to the many homonyms, many concepts have multiple ambiguous alignment candidates which need to be disambiguated in a separate step.

² <http://www.niso.org/workrooms/iso25964>

3.3 Disambiguation Techniques

In previous work we described two types of disambiguation techniques for ambiguous candidate alignments. Both techniques use the broader/narrower relationships of the source and target vocabularies (hyper/hyponym in lexical sources).

In the Child Match technique we follow, for each source concept with multiple alignments in the target vocabulary, the hierarchy “downwards” and count the number of alignments between “child” concepts of aligned concepts. We assume that concepts with similar meaning will have similar hierarchies below them. This means there should be more alignments between their children, than for homonym concepts which may be lexically similar but differ in meaning. We then count the number of alignments that have at least one or more child alignments and consider them to be correct close matches. If multiple concepts have more than one child alignment we choose the alignment with the highest number of child alignments. In some cases both alignments have the same (highest) number of child alignments, and then both alignments are chosen. The Parent Match technique is a mirroring of the Child Match technique. We find correct alignments by following the hierarchies “upward”, and count the number of aligned “parent” concepts to distinguish the correct target concept from its homonyms.

4 Case Study Setup

To answer our research questions we performed the following study. We first analyse the characteristics of the our vocabularies in our dataset. Based on this analysis we make a number of predictions about the performance on the different alignment techniques described. We then apply four different techniques and discuss the alignments sets they produce, and how these sets overlap. Lastly, we manually evaluate representative samples, and discuss the quality of the results.

4.1 Characteristics of the Vocabularies

For the two case studies we use Getty’s Art and Architecture Thesaurus (AAT)³ and its Dutch version, AATNed⁴. The two vocabularies are closely linked, in fact the AATNed was based on the AAT and extended with additional terms. We chose Princeton WordNet version 2.0 as the target vocabulary for the English language AAT and Cornetto⁵, a WordNet-like lexical resource for Dutch, as the target for AATNed.

For WordNet we used the RDF version published by W3C, the other vocabularies were originally in XML but were converted to SKOS⁶ by the MultimediaN E-Culture project. We describe them in a little more detail below:

³ http://www.getty.edu/research/conducting_research/vocabularies/aat/

⁴ <http://www.aat-ned.nl/>

⁵ <http://www2.let.vu.nl/oz/cornetto/index.html>

⁶ <http://www.w3.org/2004/02/skos/>

AAT is a structured vocabulary in English containing terms related to fine art, architecture and archival materials. It is organized in 7 facets with 36 hierarchies and contains 2,949 guide terms and 27,992 concepts. There are broader/narrower and related relations between the concepts and each has exactly one preferred label and possibly multiple alternative labels with a total of 92,089 alternative labels for concepts. Ambiguous preferred labels are distinguished from each other with the use of qualifiers. An important feature in terms of alignment is that the preferred labels are in plural form if a plural form is linguistically possible.

AATNed is a structured vocabulary in Dutch, closely related to the English AAT. It is organized in 34 hierarchies with 2,873 guideterms and 30,817 concepts. There are broader/narrower and related relations between concepts, each one having exactly one preferred label. As in the AAT, qualifiers are used to distinguish homonymous preferred labels. Concepts can also have alternative labels with a total number of 24,817 for concepts, a number significantly lower than for the AAT. As in the AAT, preferred labels tend to be in plural form. We found 20,457 singular labels for the same number of concepts meaning that 10,360 preferred labels have no singular form. These concepts tend to describe processes, states or certain materials such as “marble” or “cement” that are already in singular form or have no plural form.

WordNet is a large lexical database for the English language. It contains 115,424 synsets with 203,147 labels. A synset may contain over thirty labels and one label may appear in multiple synsets (homonyms). There are 17 relations between synsets such as the hyponym and meronym relations. Important differences with the AAT are that all labels are equivalent in the sense that there is no distinction between a preferred label and alternative label, labels tend to be in singular form and contain no diacritics.

Cornetto is a large lexical database in Dutch containing 70,370 synsets and 103,762 labels. There are 57 relations between the synsets than in WordNet, such as the “has hypernym” and “causes” relations. An important distinction between Cornetto and WordNet is that Cornetto has fewer synsets and significantly fewer labels than WordNet.

Finally, an important difference between the source vocabularies (AAT and AATNed) and the target vocabularies (WordNet and Cornetto) is that the first describe the cultural heritage domain while the latter describe more general perceptions of the world, which is often visible in the different way the hierarchies are organized. The difference in ontological commitments means that even lexical matches do not necessarily have the same meaning. One example is the concept “artist” in AAT referring exclusively to artists in the fine arts, while in WordNet the meaning also includes musical and other types of artist. As a result, finding true exact matches between the concepts is difficult, therefore our aim is to generate close matches instead. Each of the four vocabularies can be navigated through using the Europeana Thesaurus Navigator⁷.

⁷ <http://eculture.cs.vu.nl/europeana/session/thesaurus>

4.2 Predictions/Hypotheses

Based on the analysis of the characteristics of the vocabularies in our dataset, we make the following predictions.

First, AAT, AATNed and Cornetto contain diacritics in their labels, while WordNet does not. The vocabularies also differ in their use of capital letters and finally they also differ in the use of hyphens. We predict that these differences will have a significant negative effect on all alignment techniques that use simple string matching without normalization of the labels, and that this effect will outweigh the possibly negative effects of normalisation.

Second, both the AAT and AATNed contain lexical variations of their preferred label as alternative labels of the same concept. We predict that an alignment tool not restricted to preferred labels would therefore generate significantly more alignments using the lexical variations, but at the cost of lower precision.

Third, the large number of synonym labels in the target vocabularies increases the likelihood of finding alignments and therefore increasing recall. However, we also expect the precision to be low as both WordNet and Cornetto contain a large number of homonyms.

Finally, WordNet and Cornetto labels are mostly in singular form, while in AAT and AATNed there is a preference for plurals. We predict that this mismatch will have a significant negative effect on the alignment, and that this will outweigh any negative effects of errors introduced by automatically converting all terms to singular form.

4.3 Alignment Generation

We selected four alignment techniques based on the possibilities described in Section 3; three morphological techniques and a lexical technique.

Our baseline technique uses simple string-matching. It was also used in [12] in order to compare performances. It generates alignments between unique preferred labels of the source vocabularies (AAT and AATNed) and unique labels of the target vocabularies (WordNet and Cornetto). Concepts with homonymous labels are simply ignored.

The second technique matches unique singular labels. For AATNed, we use the 20,457 singular labels present in the original vocabulary. Since AAT is missing explicit singular labels, we generated them by first applying the built-in Porter stemmer of SWI Prolog⁸ to the preferred label of each concept. We then matched the resulting stem to the alternative labels of the same concept. If we found a match, we added the label as a singular preferred label to the concept. This yielded 9,129 singular labels. This is just a third of the total number of concepts, significantly less than in AATNed. The main reason for this is that the stem of a plural does not always yield the singular form, stemming for example the word “houses” to “hous”, which would subsequently not match the alternative label “house”. In addition, the Porter stemmer removes more affixes

⁸ <http://www.swi-prolog.org/>

than just the plural affixes, therefore applying it to the alternative labels could yield multiple matches to the stemmed preferred label not all of which being the singular form of the preferred label. A better approach would be the use of an algorithm that only removes plural affixes. Concepts with homonymous labels are simply ignored. We refer to this technique as the Singular Non-ambiguous technique or SN for short.

The third technique matches unique normalized singular preferred labels or normalized preferred labels if no singular label is available. These are matched to normalized labels from WordNet or Cornetto. Normalization includes replacing diacritics with a non-diacritic character (“ó” to “o”), replacing hyphens and underscores by spaces, and turning each label into lower case. Note that normalization may infrequently introduce ambiguity. For example, after normalization the Indian style “Amber” and the material “amber” have the same preferred label. Concepts with homonymous labels are simply ignored. We call this technique the Normalized Non-ambiguous technique or NN for short.

With the fourth technique, called Lexical, we match all normalized labels of the source thesauri to normalized labels of the target thesauri, regardless of whether they are unique, in order to generate as many as possible candidate alignments.

We applied all four techniques to generate mappings from AAT to WordNet, and from AATNed to Cornetto. Before applying the techniques, we removed qualifiers from the preferred labels of the AAT and AATNed as neither WordNet nor Cornetto have qualifiers. This means that we introduce ambiguity in the AAT labels, and we need to rely on the disambiguation techniques to repair this in a later phase. We then apply the Child Match and Parent Match disambiguation techniques described in Section 3 on the set of ambiguous alignments.

4.4 Manual Evaluation

Unlike in previous work where the entire set of generated alignments (4,375) were evaluated manually, we expected in these case studies to generate significantly more alignments. Evaluating a large number of alignments manually is not feasible. We sampled alignments from various subsets we expected to have different properties. These samples were evaluated manually. We also performed inter-rater agreement evaluations to check the quality of the manual evaluation. Subsequently, we extrapolate from the results of the evaluated samples to estimate the precision of the subsets using the method described in Van Hage et. al. [9].

4.5 Summary of case study setup

In **Step 1** we preprocess the data-sets by removing qualifiers from the preferred labels. In **Step 2** we apply the four alignment techniques to the vocabularies to generate close-match alignments. In **Step 3** we apply two disambiguation techniques that use the structure of the vocabularies. In **Step 4** we perform manual evaluation of samples of data classifying each alignment into one of

Table 1. Number of alignments generated between AAT and WordNet

Method	Alignments	Aligned AAT Concepts	% of AAT Concepts
Baseline	2,296	2,296	8%
SN	4,299	4,299	15%
NN	4,365	4,365	15.5%
Lexical	42,039	12,725	45%
Total	42,039	12,725	45%

Table 2. Disambiguated alignments of the Lexical tool for AAT-WordNet

Segment	Kept Alignments	Disambiguated Concepts	Removed Alignments
Child Match only	590	554	3,205
Parent Match only	2,485	2,011	7,035
Overlap	236	234	1,580
Distinct Total	3,311	2,665	11,820

seven categories: exact-match, close-match, broader, narrower, related, unsure and not related. We also record the amount of time the evaluation takes. In **Step 5** independent raters evaluate random samples of evaluated alignments in order to calculate inter-rater agreement statistics (Cohen’s Kappa). In **Step 6** we estimate based on the results of the evaluated alignments the performance of the alignment and disambiguation techniques. The focus here is on alignments evaluated as exact-match and close-match.

5 Alignment Results

5.1 AAT-WordNet

Alignment Generation We generated four sets of alignments using the four alignment techniques. Table 1 displays the number of alignments per technique with the number of AAT concepts aligned and the percentage of the total AAT concepts. The baseline generates the fewest alignments which was expected, this is caused by the large number of preferred labels in plural form. The SN technique is more successful aligning almost twice as many concepts. The NN tool only generates 66 alignments more than the SN tool. Combined, the three morphological tools generate 4,592 distinct alignments which is 16.4% of the AAT concepts. The Lexical tool generates almost ten times more alignments for three times the amount of concepts the morphological tools generate. This is caused by the large number of alternative labels of the AAT and allowing for ambiguous alignments.

Combining Alignments Examining the overlap between the alignments generated by the four tools revealed that all alignments generated by the morphological tools were also generated by the Lexical tool. Fig. 1 shows the overlaps between the three morphological tools. There is a large overlap between all three tools along with a large overlap between the SN and NN tools. The figure also

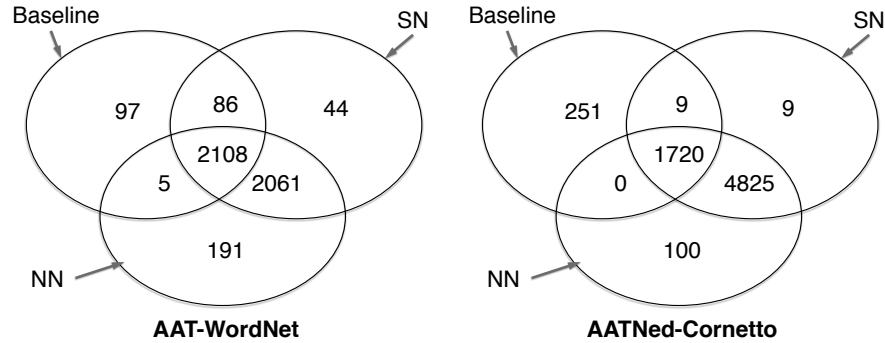


Fig. 1. Overlap in alignments generated by the morphological tools

shows that 191 NN alignments were *not* found by the Baseline and SN tools. Most of these alignments are upper case labels matched to lower case labels or normalized diacritics e.g.: “venetian blind” matching “Venetian blind”. An example of an alignment only found by SN is “Maltese cross”. When normalized the term matches two synsets in WordNet and becomes ambiguous. The Lexical tool generated an 37,447 alignments for 8,990 concepts excluding alignments displayed in Table 2. Further analysis showed that 2,116 of these alignments were not ambiguous. These non-ambiguous alignments are between unique AAT alternative labels and unique WordNet labels.

Disambiguation Of the 12,725 concepts that were aligned, 6,887 concepts have more than one alignment with a total of 36,201 alignments. We applied Child Match and Parent Match disambiguation techniques where we rely on the structure of the thesauri. Analysis of the results of the disambiguations showed that 908 alignments were also generated by the morphological tools. We removed these alignments as our goal is to improve the performance of the Lexical tool.

Table 2 displays the results of the disambiguation process showing the alignments that were kept, the number of disambiguated concepts and the alignments that were rejected. The Parent Match technique disambiguated three and a half times more concepts than the Child Match technique. There is a small overlap between the two. Examples of correctly disambiguated alignments are the concept “vehicle” as in motorized vehicle which is disambiguated from “vehicle” meaning expression or medium because of alignments between its children such as “aircraft” and “tricycle”. Of the 6,887 concepts, 2,665 concepts (39%) were disambiguated.

By combining the non-ambiguous alignments found by the four techniques with the results of the disambiguation techniques we have 10,0019 alignments for 9,208 AAT concepts.

5.2 AATNed-Cornetto

Table 3. Alignments generated between AATNed and Cornetto

Method	Alignments	Aligned AATNed Concepts	% of AAT Concepts
Baseline	1,980	1,980	6.5%
SN	6,563	6,563	21%
NN	6,644	6,644	22%
Lexical	20,331	10,773	35%
Total	20,331	10,773	35%

Table 4. Disambiguated alignments for AATNed-Cornetto

Segment	Kept Alignments	Disambiguated Concepts	Removed Alignments
Child Match only	342	327	1,140
Parent Match only	1,281	920	1,667
Overlap	106	104	289
Distinct Total	1,729	1,297	3,096

Alignment Generation The result of the alignment process is shown in Table 3. Similarly to AAT-WordNet case, Baseline generated the fewest number of alignments. The SN tool generated over three times that many alignments as more singular labels were available. In total, the three morphological techniques generated 6,923 alignments for the same amount of concepts, aligning a little over 22% of AATNed concepts.

The Lexical tool generated 20,331 alignments for over a third of the total AATNed concepts and there are significantly fewer alignments generated than for AAT-WordNet. This is caused by the lower number of alternative labels in AATNed and fewer sense labels in Cornetto.

Combining Alignments Again, all alignments generated by the three morphological tools were also generated by the Lexical tool. Fig. 1 shows the overlap between the non-ambiguous tools. There is a large overlap between the three tools. However, the number of alignments found only by Baseline tool is larger than in the English case. There is an even larger overlap between SN and NN tools.

In addition to the alignments shown in Fig. 1, the Lexical tool generated 13,417 alignments for 4,414 concepts. An analysis showed that a small subset of these alignments (569) is not ambiguous. This is smaller number than in AAT-WordNet, again caused by the fewer alternative labels.

Disambiguation Of the 10,773 concepts that were aligned in total, 3,899 concepts have more than one alignment with a total of 13,457 alignments. We applied Child Match and Parent Match and removed 142 alignments also generated by the morphological tools. The results are shown in Table 4. Again, there is a small overlap between the two techniques. Overall 1,297 concepts were disambiguated, which is a third of the total number of ambiguous concepts. In comparison to the AAT- WordNet case we see that a smaller percentage of the

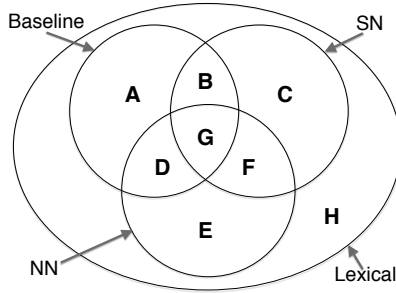


Fig. 2. Venn diagram of segments representing the overlaps of the four tools

aligned concepts are ambiguous. This is caused by fewer number of alternative labels in AATNed and also fewer labels per concept in Cornetto.

6 Evaluation

We manually evaluated samples of generated alignments. The samples were selected from all segments of the Venn diagram shown in Fig. 2. For segment H, we the samples are from alignments returned by the two disambiguation techniques and their overlap, as well as alignments removed by the two techniques in order to assess calculate the number of false negatives. We also took a sample of ambiguous alignments that were not disambiguated to assess the overall performance of the Lexical tool. Finally, we also sampled non-ambiguous alignments generated using unique alternative labels. In total we selected alignments for one thousand source concepts from AAT and one thousand concepts from AATNed. These alignments were scattered along all segments: 400 concepts were sampled segments A to F, and 600 concepts from segment H.

The alignments were manually evaluated by the first author using the evaluation tool used in [12]. Three extra raters evaluated small samples of the evaluated set to measure inter-rater agreement. Each rater had a different sample of 50 concepts from with alignments of both cases. Averaged over the three raters Cohen’s kappa was 0.67 (for English) and 0.71 (for Dutch) which is a moderate agreement. This result shows that alignment evaluation is difficult, even for humans.

6.1 Results for AAT-WordNet

Table 5 displays the result of the manual evaluation. In the columns for the AAT-WordNet alignments we see that the overlap between the SN and NN tools marked by segment F has the highest precision (0.9), followed by the segment G at 0.81. As predicted, the precision of the alignments generated by the Baseline tool only in segment A, is lower than most other segments at 0.38. This is caused by plural nouns from AAT incorrectly matching verbs in WordNet.

Table 5. Sampled evaluation from the non-ambiguous segments

Segment	AAT-WordNet				AATNed-Cornetto			
	Segment Size	Sample Size	Correct Alignments	Sample precision	Segment Size	Sample Size	Correct Alignments	Sample precision
A	97	55	21	0.38	251	132	66	0.5
B	86	46	20	0.43	9	9	4	0.44
C	44	44	35	0.79	9	9	5	0.56
D	5	5	5	1.0	0	0	0	0
E	191	50	18	0.36	100	50	24	0.48
F	2061	100	90	0.90	4825	100	92	0.92
G	2108	100	81	0.81	1720	100	87	0.87

Table 6. Sampled evaluation of alignments found only by the Lexical technique

Segment	AAT-WordNet				AATNed-Cornetto			
	Segment Size	Sample Size	Correct Alignments	Sample precision	Segment Size	Sample Size	Correct Alignments	Sample precision
ChildMatch only	590	101	75	0.74	342	88	78	0.89
ParentMatch only	2485	120	55	0.46	1281	135	92	0.68
Child Match Parent Match overlap	236	44	37	0.84	106	47	43	0.91
Discarded Child-Match only	3205	226	17	0.08	1140	161	19	0.12
Discarded Parent-Match only	7035	178	11	0.06	1667	147	9	0.06
Discarded Child-Match ParentMatch overlap	1580	119	4	0.03	289	61	5	0.08
Remaining ambiguous alignments	20200	545	62	0.12	8023	306	59	0.19
Lexical only unambiguous al.	2116	100	65	0.65	569	100	74	0.74

Table 6 shows the results of the sampling of alignments from the set found only by the Lexical tool. The disambiguation with the Child Match technique performs better with a precision of 0.74 while the ParentMatch technique has a precision of 0.46 and the overlap the highest precision at 0.84. The number of false negatives in the discarded segments is low at between 3% and 8%. Looking at the remaining ambiguous alignments that were not disambiguated we have a precision of 0.12. This means from the 20200 ambiguous alignments an estimated 2650 alignments should be close- or exact-matches. The non-ambiguous alignments generated using alternative labels have a precision of 0.65. This is lower than for alignments with preferred labels, supporting the view that alternative labels yield worse alignments than preferred labels. We estimate the precision of all the alignments between AAT and WordNet without disambiguation at 0.17. Thus, only applying lexical alignment without disambiguation yields an unacceptably low precision.

6.2 Results for AATNed-Cornetto

The results of the evaluation displayed in Table 5 and Table 6 show that the overall precision of the techniques is higher for the Dutch than for the English

Table 7. Precision and coverage of combined alignment techniques

Segment	AAT-WordNet				AATNed-Cornetto			
	Alignm.	Conc- epts	Prec.	% AAT Concepts	Alignm.	Conc- epts	Prec.	% AATNed Con- cepts
1. Union morph. tools	4,592	4,592	0.82	16.4%	6,914	6,914	0.88	22.4%
2. non-amb. Lexical	2,116	2,116	0.65	7.6%	569	559	0.74	1.8%
3. Disam- biguated align- ments	3,311	2,665	0.53	9.5%	1,729	1,297	0.74	4.2%
1 + 2	6,708	6,695	0.70	23.9%	7,483	7,425	0.86	24.1%
1 + 2 + 3	10019	9,208	0.69	32.9%	9,212	8,621	0.84	27.9%

language vocabularies. This is caused by the lower number of labels per concept in both AATNed and Cornetto resulting in fewer alignments per concept.

Similarly to the results of AAT-WordNet, half of the sample alignments generated only by the Baseline tool, (segment A) are incorrect. There were more alignments in this segment than in the AAT-WordNet and the evaluation revealed that 60% of these alignments were to Cornetto verbs. Some of the alignments are correct. The labels of processes in AATNED, for example, are often verbs. However, most others matches to verb targets are incorrect. For example, the concept “handwerken” (needle-works) is mapped incorrectly to the verb “handwerken” (needle-working). Again, just as for the AAT-WordNet, the alignments generated by the NN only (segment E) have relatively low precision at 0.48, although this is higher than the 0.36 for AAT-WordNet segment. Most of the erroneous alignments are due to concepts that describe styles and periods aligned to the nationality or language the style gets its name from. These concepts are related but are not the same (e.g. the “Pueblo” style and “pueblo (house)”). The overlap between SN and NN (segment F) has the highest precision, following the trend we have seen in AAT-WordNet.

Table 6 shows that the disambiguation techniques performed slightly better in Dutch than in English. Although difficult to analyze, this is possibly due to more similar hierarchies in the Dutch vocabularies than the English vocabularies. The overlap between the two techniques has the highest precision of 0.91 followed by the alignments found only by the Child Match technique at 0.89 and the Parent Match at 0.68. The number of false negatives ranges from 6% to 12%. The precision of the sample of the subset of 569 non-ambiguous alignments generated by the Lexical tool is 0.74. Finally, we estimated the precision of all alignments between AATNed and Cornetto at 0.46. This is significantly higher than in the AAT-WordNet case, caused by fewer alternative labels in the AATNed and the fewer labels per concept in Cornetto.

6.3 Results of Alignment Technique Combination

We now look at the precision of combined techniques for AAT-WordNet and AATNed-Cornetto. Table 7 displays the number of alignments, concepts, their precision and the percentage of all concepts they represent. The union of morphological tools combined with non-ambiguous Lexical alignments (of alternative labels) align 24% of AAT and AATNed concepts with precision of 0.7 and 0.86 respectively. By further adding the disambiguated alignments from the Lexical tool the coverage in terms of aligned concepts increases to 32.9% for the English case and 27.9% for the Dutch case although precision drops slightly to 0.69 and 0.84. However, this seems to be an acceptable trade-off for the boost in coverage.

7 Conclusion

We make the following three conclusions about the performance of the techniques.

1. The simple non-ambiguous morphological matching techniques work well with a high precision but low coverage.
2. The lexical matching technique using alternative labels improves coverage but reduces precision
3. Disambiguation of the lexical matches increases coverage keeping precision at an acceptable level.

Our key findings with respect to the characteristics of the vocabularies are: First, language does not seem to be a factor in the alignment results. All differences can be explained in terms of differences in the vocabulary characteristics such as the difference in the number of alternative labels.

Second, the selection of alignment techniques is mainly influenced by the characteristics of the vocabularies. For example, in this case the use of plural labels in AAT made it necessary to extract singular labels while this was not the case for AATNed nor in the previous case study where all labels were in singular form.

Finally, the number of aligned source concepts is influenced by the difference in domain of the vocabularies, the source vocabularies being specialist cultural heritage vocabularies and the target vocabularies covering a “common sense” domain.

We found that a combination of morphological and lexical alignment techniques with disambiguation works relatively well given the differences in vocabularies. For future work, using additional background knowledge for example about the partitioning of vocabularies with regard to part-of-speeches is expected to further increase precision and coverage.

Acknowledgement The datasets have been kindly provided by RKD and the Cornetto project. We thank Marieke van Erp for her contributions on the alignment evaluation. This research was supported by the MultimediaN project funded through the BSIK programme of the Dutch Government.

References

1. Aleksovski, Z., van Hage, W.R., Isaac, A.: A Survey and Categorization of Ontology-Matching Cases. In: Shvaiko, P., Euzenat, J., Giunchiglia, F., He, B. (eds) *Proceedings of the Workshop on Ontology Matching (OM2007) at ISWC/ASWC 2007* (2007)
2. Caracciolo, C., Euzenat, J., Hollink, L., Ichise, R., Isaac, A., Malaisé, V., Meilicke, C., Pane, J., Shvaiko, P., Stuckenschmidt, H., Sváb-Zamazal, O., Svátek, V.: Results of the Ontology Alignment Evaluation Initiative 2008. In: *OM 2008* (2008)
3. Eckert, K., Meilicke, C., Stuckenschmidt, H.: Improving Ontology Matching Using Meta-Level Learning. In: Aroyo, L., Traverso, P., Ciravegna, F., Cimiano, P., Heath, T., Hyvnen, E., Mizoguchi, R., Oren, E., Sabou, M., Simperl, E.P.B. (eds) *ESWC 2009* p.158–172 (2009)
4. Euzenat, J., Ehrig, M., Jentzsch, A., Mochol, M., Shvaiko, P.: Case-Based Recommendation of Matching Tools and Techniques. deliverable 1.2.2.2.1, Knowledge Web NoE, 2006 (2006)
5. Euzenat, J., Ferrara, A., Hollink, L., Isaac, A., Joslyn, C., Malaisé, V., Meilicke, C., Nikolov, A., Pane, J., Sabou, M., Scharffe, F., Shvaiko, P., Spiliopoulos, V., Stuckenschmidt, H., Sváb-Zamazal, O., Svátek, V., dos Santos, C.T., Vouros, G.A., Wang, S.: Results of the Ontology Alignment Evaluation Initiative 2009. In: *OM 2009*(2009)
6. Euzenat, J., Isaac, A., Meilicke, C., Shvaiko, P., Stuckenschmidt, H., Sváb, O., Svátek, V., van Hage, W.R., Yatskevich, M.: Results of the Ontology Alignment Evaluation Initiative 2007. In *OM 2007* (2007)
7. Euzenat, J. Shvaiko P.: *Ontology Matching*. Springer, Heidelberg (2007)
8. Ghazvinian, A., Noy, N.F., Musen, M.A.: Creating Mappings for Ontologies in Biomedicine: Simple Methods Work. In: *2009 AMIA Annual Symposium* (2009).
9. van Hage, W.R., Isaac, A., Aleksovski, Z.: Sample Evaluation of Ontology-Matching Systems. In: *Proceedings of the ISWC workshop on Evaluation of Ontologies and Ontology-based tools* p. 41–50 (2007)
10. Hu, W. Qu Y.: Falcon-AO: A Practical Ontology Matching System. *J. Web Semant.* 6(3):237–239 (2008)
11. Schreiber, G., Amin, A., Aroyo, L., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Omelayenko, B., van Ossenbruggen, J., Tordai, A., Wielemaker, J., Wielinga, B.: Semantic Annotation and Search of Cultural-Heritage Collections: The Multimedial E-Culture Demonstrator. *J. Web Semant.* 6(4):243–249 (2008)
12. Tordai, A., van Ossenbruggen, J., Schreiber, G.: Combining Vocabulary Alignment Techniques. In: *K-CAP '09: Proceedings of the Fifth International Conference on Knowledge Capture* p. 25–32 (2009)
13. Zhang, X., Zhong, Q., Li, J., Tang, J.: RiMOM Results for OAEI 2008. In: Shvaiko, P., Euzenat, J., Giunchiglia, F., Stuckenschmidt, H. (eds) *OM* (2008)